99

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/629,175	07/31/2000	Ophir Frieder	7519-164345	4562

7590 08/04/2006

Staas & Halsey LLP
1201 New York Avenue N W
Suite 700
Washington, DC 20005

EXAMINER	
LE, UYEN T	
ART UNIT	**PAPER NUMBER**
2163	

DATE MAILED: 08/04/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 09/629,175
Filing Date: July 31, 2000
Appellant(s): FRIEDER ET AL.

Mark J. Henry
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 27 June 2006 appealing from the Office action mailed 9 September 2005.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

The amendment after final rejection filed on 14 November 2005 has not been entered because claim1 raises new issues that would require further consideration and search.

The amendment after final rejection filed on 6 April 2006 has been entered because it simply cancels claim 29.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct. It is noted that applicant did not argue each independent claim 50, 51 separately. Claim 51 recites means plus functions in "means for obtaining a document". Since claim 50 recites a corresponding method and appellant points to the structure, material and acts described in the specification as corresponding to the claimed function, the examiner consider the summary of the claimed subject matter compliant.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is correct.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

6,240,409	AIKEN	5-2001
5,136,646	HABER et al	8-1992

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Objections

Claim 45 is objected to under 37 CFR 1.75(c), as being of improper dependent form for failing to further limit the subject matter of a previous claim. Applicant is required to cancel the claim(s), or amend the claim(s) to place the claim(s) in proper dependent form, or rewrite the claim(s) in independent form. Claim 1 from which claim 45 depends already includes filtering based on parts of speech.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-28, 45-48, 50-57 are rejected under 35 U.S.C. 103(a) as being unpatentable over Aiken (US 6,240,409) of record.

Regarding claims 1, 45, Aiken discloses a method for detecting similar documents including all the claimed subject matter (see Figures 1a, b, column 3 lines 44-47). Note the step of obtaining a document 102, filtering the document 106. The claimed step of generating a tuple for the filtered document is met by the fact that a hash value and position pair is created and stored (see step 114, column 6, lines 7-28). The tuple is clearly compared with a plurality of tuples as claimed. Aiken discloses detecting if the document is similar to another document by determining if the tuple is clustered with another tuple in the document storage structured (see Figures 4a, 4b, 4c, column 7, lines 25-34, column 10, line 4- column 12, line 2). The claimed "tokens being eliminated based on parts of speech" is met by the fact that the method of Aiken eliminates stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3). Although Aiken does not specifically show sorting the filtered document to reorder the tokens according to a predetermined ranking, official notice is taken that it is well known in the art that different operating systems use different tokens ordering. Therefore, it would have been obvious to one of ordinary skill in the art to include sorting the filtered document to reorder the tokens according to a predetermined ranking in order to accommodate different operating systems while implementing the method of Aiken.

Regarding claim 2, Aiken discloses parsing and filtering the document (see column 4, lines 54-67). Clearly the filtered document comprises a token stream of a plurality of tokens as claimed.

Regarding claim 3, Aiken discloses retaining a token according to at least a token threshold (see column 11, lines 15-30) and tokens frequently .

Regarding claim 4, Aiken discloses that the retained tokens are arranged in the token stream (see Figure 4a, step 404).

Regarding claim 5, Aiken discloses determining the hash value for the filtered document by processing individually each retained token in the token stream (see column 6, lines 7-28, column 9, lines 24-26).

Regarding claim 6, Aiken discloses determining a score for each token in the token stream and comparing the score for each token to a first token threshold (see column 11, lines 15-30). The token stream is clearly modified by removing each token having a score not satisfying the first token threshold and retaining each token having a score satisfying the first token threshold as claimed since the document not containing a certain match ratio is discarded in the method of Aiken.

Regarding claim 7, although Aiken does not specifically show the step of comparing the score for each retained token to a second token threshold and modifying the token stream as claimed, Aiken explicitly show that not every substring's hash value is stored (see column 6, lines 29-30). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed feature while implementing the method taught by Aiken in order to further filter the document and save memory.

Regarding claim 8, Aiken discloses filtering by removing from the token stream at least one token corresponding to a stop word (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 9, although Aiken does not explicitly disclose filtering by removing a duplicate of another token in the token stream, it would have been obvious to one of ordinary skill in the art to include such a feature in order to avoid processing redundant token, thus saving time and resources.

Regarding claim 10, Aiken discloses removing a token from a token stream if the token is a very frequent token when Aiken shows that the method remove words of "the" "and" , "this", "is" (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 11, Aiken discloses removing a token from a token stream (see column 4, lines 57-58, column 8, line 67- column 9, line 3).

Regarding claim 12, Aiken discloses removing formatting from the document (see column 4, lines 55-57).

Regarding claims 13, 14, clearly the method of Aiken uses collection statistics pertaining to a plurality of documents for filtering the document since the input file is compared to a set of collected files to detect similarity (see column 2, lines 47-51). The collection statistics have to be present for the collected documents to be clustered as shown in the method of Aiken (see Figure 4c, column 11, line 47- column 12, line 2).

Regarding claims 15-18, although Aiken does not explicitly show that the method uses specific hash algorithms as claimed, it is notoriously well known in the art to use different hash algorithms depending on users' requirements. Therefore, it would have been obvious to one of ordinary skill in the art to include all the claimed features while implementing the method of Aiken in order to suit users' needs.

Regarding claim 19, Aiken discloses a hash table (see column 12, lines 40-44).

Regarding claim 20, Aiken discloses that the document storage structure comprises a tree (see column 8, lines 30-38).

Regarding claims 21, 22, Aiken discloses that the tree comprises a binary tree (see column 8, lines 36-38). Although Aiken does not explicitly show that the binary tree is balanced, it would have been obvious to one of ordinary skill in the art to include such a feature in order to store data efficiently and to facilitate searching and localization.

Regarding claim 23, Aiken discloses a hash table and at least one tree (see column 5, lines 33-40, column 8, lines 30-38).

Regarding claim 24, Aiken discloses inserting the tuple into the document storage structure (see Figure 1a, 1b, 4a, 4b, 4c).

Regarding claim 25, the hash table of Aiken clearly comprises a plurality of bins of tuples as claimed and the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple at a bin of a hash table (see Figures 1, 2, 4c, column 7, line 46- column 8, line 33).

Regarding claim 26, Aiken discloses a tree comprising a plurality of branches, each bucket of the tree comprising at least one tuple and wherein the step of determining if the tuple is clustered with another tuple clearly comprise determining if the tuple is co-located with another tuple in a bucket of the tree (see column 8, lines 31-54, Figure 4c).

Claim 27 corresponds to a system to perform the method of claim 1, thus is rejected for the same reasons stated in claim 1 above.

Claim 28 corresponds to a computer program product to perform the method of claim 1, thus is rejected for the same reasons stated in claim 1 above.

Regarding claim 46, Aiken discloses removing frequently occurring terms (see column 4, lines 48-53).

Regarding claims 47-48, although Aiken does not specifically show removing infrequently occurring terms or words having an occurrence frequency that falls within a pre-determined frequency range, since users requirements vary, it would have been obvious to one of ordinary skill in the art to include the claimed features in order to accommodate users applications.

Regarding claim 49, although Aiken does not specifically show Unicode ordering, since Unicode is a recognized standard, it would have been obvious to one of ordinary skill in the art to include such ordering in order to use a standardized technique while implementing the method of Aiken.

Claim 50 recites the limitations of claim 1 without the sorting step, thus is broader than claim 1 and is rejected for the same reasons stated in claim 1 above.

Claim 51 corresponds to a system for claim 50, thus is rejected for the same reasons stated in claim 50.

Regarding claims 52-57, the claimed criteria for determining threshold and frequency scores merely read on notoriously well-known decision making techniques in the art. Therefore, it would have been obvious to one of ordinary skill in the art to include any criteria deemed appropriate while implementing the method of Aiken depending on users requirements.

Claim 44 is rejected under 35 U.S.C. 103(a) as being unpatentable over Aiken (US 6,240,409) of record, further in view of Haber et al (US 5,136,646) of record.

Regarding claim 44, Aiken discloses determining a hash value for a document (see Figure 1, column 4, line 17- column 7, line 45, column 9, lines 16-30), accessing a document storage structure comprising a plurality of hash values, each hash value representing one of a plurality of documents (see Figure 4a, column 10, line 4- column 11, line 46), determining if the hash value is equivalent to another hash value in the document storage structure (see Figure 4c, column 11, line 47- column 12, line 2). Although Aiken does not specifically show each tuple comprises a document identifier and a single hash value, it is well known in the art to hash a document into a single hash value as shown by Haber (see the abstract). Therefore, it would have been obvious to one of ordinary skill in the art to include the claimed features while implementing the method of Aiken in order to detect document similarity instead of just portions of a document.

(10) Response to Argument

Appellant argues at page 13 that

"independent claims 50 recites filtering the document to eliminate tokens based
"on parts of speech, independent claim 51 recites a filter to filter the document to
"eliminate tokens based on parts of speech, independent claim 1 recites tokens
"being eliminated based on at least one of (a) parts of speech and (b) collection
"statistics. The examiner failed to give the words in the claims the broadest

"meaning such terms would allow. The examiner failed to consider grammatical "parts of "speech by equating parts of speech with stop words.

In response to the argument that the examiner failed to give the words in the claims the broadest meaning such terms would allow, the examiner respectfully disagrees. The claim language does not require filtering based on all parts of speech of a spoken language, thus interpreted broadly, any part of speech for example article, preposition that happen to be stop words would meet the limitation of "filtering based on parts of speech".

In response to the argument that the examiner failed to consider grammatical parts of speech by equating "parts of speech" with stop words, the examiner respectfully disagrees. First, the claim language does not include "grammatical". Second, although some words in a spoken language are considered stop word, all words must be parts of speech for example "the", "a", "an" are grammatically articles, "of', "to" or "for" are grammatically prepositions. The examiner recognizes that not all parts of speech are prepositions or articles but all prepositions or articles are clearly parts of speech. Appellant even admitted that "parts of speech" includes nouns, verbs, adjectives, prepositions, and types of nouns at page 13, 5th paragraph. Thus, the examiner maintains that the stop words and words frequently used in Aiken read on the parts of speech of claims 1, 50 and 51. Aiken in an example also shows stop words "this" and "is" in Figure 3 and column 9, lines 1-3. The words "this" and "is" are clearly grammatical parts of speech because "this" is a pronoun and "is" is a verb.

Appellant alleges that the present invention explicitly distinguishes token removal based on parts of speech and token removal based on stop words.

In response, the examiner respectfully disagrees. Claim 8 that recites "removing from the token stream at least one token corresponding to a stop word" clearly shows no distinction between a stop word and a part of speech.

Applicant argues that the examiner did not explicitly address the features of claim 1 of "eliminated based on at least one of (a) and (b)".

In response, claim 1 does not require both limitations (a) and (b), only at least one of the two. Therefore, the examiner did not address the limitation in (b). Besides, limitation (b) does not require the collection statistics to be relating to a number of occurrences of all words or phrases in the document. Thus, the number of occurrences of frequent words in Aiken reads on the claimed collection statistics relating to a number of occurrences of words or phrase in the document.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

26 July 2006

Uyen Le

Ule

UYEN LE
PRIMARY EXAMINER

Conferees:

Don Wong

DON WONG
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

Jeffrey Gaffin